**COVER LETTER**

**Submission ID**: 128

**Title**: Interactive Exploration of Ligand Transportation through Protein Tunnels

**Date**: July 17th, 2016

We would like to thank the anonymous reviewers for their helpful comments. We attempted to address all remarks in the revised version of the paper. In this letter, we give detailed responses to all reviewer comments and describe the corresponding changes in our manuscript.

**Reviewer 3 - primary**

1. *There was one significant scientific concern raised by several reviewers, and ask that you do pay attention to that: specifically your choice of trajectory simplifications raises questions regarding its properties and suitability.  There is previous work on trajectory simplification, some quite well studied.  Two of the reviewers pointed out specific works that might be appropriate to consider. It would be helpful to your readers to at least put your algorithm in the context of previous (molecular) trajectory simplification work, or even to provide a choice of simplification algorithms, there are interesting differential benefits between them.*

When developing the system, we were experimenting with both smoothing and non-smoothing simplification approaches. We expected that we will be able to automatically classify parts of a smoothed trajectory, but the expectation was not confirmed. Therefore, by leaving the classification idea, we were able to provide the domain experts also with non-smoothing simplification based on the Douglas-Peucker algorithm. The domain experts concluded that this additional approach is beneficial for them due to its nature, i.e., keeping points from the original trajectory.

We added a brief description about the non-smoothing approach to the end of the Section Automatic Simplification.

1. *From a vis perspective it could use some more work evaluating the correctness of what has been done, and whether there might be alternative solutions that are superior, but, from a biological perspective it's clear that the approach used in the manuscript "works", and having something that works out there for people who are trying to understand molecular ligand/docking trajectories, would be a real plus.*

The correctness of our solution was evaluated namely from the biochemical point of view. We aimed to explore if the tool is beneficial for the biochemists and if it gives them the valuable insight to the data. During the design phase we explored different potential visual representations and their applicability to our problem, which due to the space limitations were not included. Finally, according to the acquired feedback we concluded that the selected set of visualizations is the most appropriate and understandable by the domain experts.

1. *One area where I wish the manuscript contained more detail and comparison, is in the simplified trajectory material.  It appears that the simplification mechanism "works", but it's not obvious how well it works, or how robust it is to small variations in the input.  For example, it would be nice to know that the simplification produced from one set of MD runs for a ligand into a pocket, is similar to the simplification produced from a different set of MD runs for the same ligand and pocket.  It would also be nice to have more evidence that the simplification reliably preserves "interesting features" in the trajectory.  Admittedly, there are some challenges to performing these studies, but, whether in this paper or a follow-up, it would be good to give the reader some more information in this area.*

To address the robustness issue, we performed an experiment in order to evaluate the stability of the trajectory simplification. We used a simulation of molecular dynamics and its artificially modified counterpart. In order to obtain the artificial MD, we moved every ligand position by a random unit vector. Then, we applied the simplification to both trajectories and we compared the results at several levels of magnitude. We concluded that the simplification is stable wrt. perturbations along different MD runs.

To support our conclusion, we add figures depicting results of our experiment for four different levels of magnitude.

|  |  |
| --- | --- |
| Original MD | Artificial MD |
|  |  |
|  |  |
|  |  |
|  |  |

Regarding the evaluation of “interesting features” preservation, the domain experts were from the beginning interested in parts of the trajectory where the ligand can pass “quickly”. These parts were hard to follow in 3D when visualizing the whole trajectory. Therefore, we addressed the issue using simplification which preserved these “quick pass” parts as their trajectories are “simple”. Moreover, we obtained overviews of “complex” parts that also can be directly communicated using 3D visualization.

Finally, we considered incorporating a second case study into the paper rather than including the ideas presented above, which nevertheless should be elaborated in next studies.

**Reviewer 1**

1. *There are some minor questions on the design choices (see below). My main question would be about the feedback. Apparently, the biochemists have used the tool. By themselves? What was the major gain? New insight? Or just more efficient? By how much did the efficiency increase?*

As we aimed to design the tool to fit the biochemical needs, the cooperating group of biochemists was directly involved into the design phase. In the evaluation phase the biochemists were asked to load their datasets of interest and to evaluate them. To cover this, we added the following text to the Analysis Procedure and Discussion Section:

“These scenarios were selected and conducted by the domain experts from our cooperating group of protein engineers. The group involved into the design of our proposed tool as well as the selection of interesting case studies and evaluation of the final visualization consisted of seven researchers - one professor (head of the protein engineering group), two post-docs, and four PhD students. They were asked to use our tool to explore their datasets and to evaluate the intuitiveness, understandability, benefits, and drawbacks of the proposed tool.”

1. *The simplification of the trajectories may not necessarily be the best choice. I would have opted for a multiresolution approach, where small details are successively omitted, thus effectively applying a low-pass filtering. Then, a simple slider could be used to intuitively and interactively choose the amount of simplification desired.*

The majority of discussions with the biochemists during the design phase was related to the selection of the best simplification method. There are plenty of possible approaches in the literature which could be adopted and could be potentially beneficial. We also tried to implement some of them and finally agreed on the approach described in the paper. However, as we also concluded that there is no ideal solution to this problem, and we decided to combine the automatic approach with the interactive one so the user can decide on the level of simplification in different parts of the trajectory. The interactive simplification in fact supports the manual setting of the amount of simplification using a slider.

1. *Another parameter is the window chosen for smoothing the scatterplots. Here also the question is how much is desired and whether the control should be given to the user.*

As the paragraph Scatterplot Matrix in Section Visual Exploration describes, the need for this parameter stems from the noise in the raw data caused by the ligand jittering. This noise often obscured relationship and trends between attributes displayed in Scatterplot Matrix. The sliding window parameter, i.e., the smoothing functionality, was explicitly requested by the domain experts and helps to eliminate this problem. However, the exploration of data at multiple levels of detail is still often required, particularly so in combination with selection and zooming. Therefore, we provide interactive control of the sliding window parameter, which enables quick way to smooth the data to desired level.

1. *The analysis starts with an overview of the entire trajectory. This aspect does not scale to very large number of time steps. Any thoughts on this?*

The overview representation was required and highly appreciated by the domain experts. It gives them the idea about the overall behavior of the ligand which is almost impossible to capture from the original data. As for the scalability, we are able to compute the overview for simulation consisting of 50.000 time steps in order of minutes. As such, the overview immensely contributes to the ability to explore the long dynamic in reasonable time.

1. *The Direction parameter sounds like a binary value. I assume it is more like the derivative of the distance, which would make a lot of sense, but it is not so clear from the description.*

We corrected the description of the Direction parameter in the following way (described in Section Derivation of Attributes):

“The direction parameter is a binary value computed as the derivative of the distance attribute. In other words, we simply evaluate the ligand distance from the active site in two subsequent time steps and if the difference of the obtained values is positive we claim that ligand is moving towards active site and vice versa.”

1. *I wonder about the color choices in the overview visualization. Why were the colors chosen as they are. There is no obvious intuitive interpretation for me and green and blue are hard to distinguish. I actually don't see any blue on my print-out.*

At first, we intended to choose such colors which would be dissimilar to the color schemes used for encoding the physico-chemical properties of amino acids. Thus we have chosen pastel shaded colors. After further examination of multiple cases, this scheme is completely inapplicable. We would like to thank to the reviewer for this valuable comment.

In this second review, we have chosen a diverging and colorblind safe color scheme that is visually leading the user to the interesting areas (when is the ligand at the active site). This color scheme is slightly similar to the encoding of partial charge of amino acids. Nevertheless, we do not consider this to be a critical problem since the visualization supports the differentiation of the context in both cases.

1. *The description of the coloring of the line charts also remained unclear to me.*

The line representation showing ligand-lining amino acids is colored according to the selected property of adjacent amino acids. We added the used color schemes into Figure 7 since we had no free space to describe it in the text. The color schemes for the properties were given by the biochemists and are consistently used in the whole CAVER Analyst tool which contains the implementation of our system.

**Reviewer 2**

1. *This is well written paper and on a specific domain problem. I really don't have any major comments but I'd encourage authors to make their tool and code publicly available for reproducibility and impact.*

The tool will be publicly available via next version of CAVER Analyst (downloadable at www.caver.cz). Before its release, we will be happy to provide the alpha version of the tool on demand.

Minor:

1. *Why not use focus+context line graphs as opposed to scatterplots for visualizing attributes along paths. Using scatterplots seems a strange choice, particularly after trajectories are smoothed.*

We discussed this issue and we concluded that line graphs would be appropriate in cases when one of the depicted attributes is shown over time (or any other attribute with linear progression). For another attributes, for instance, Hydrophobicity vs. Speed, the scatterplot is a natural choice here.

1. *The way the terms manual and automatic are used is a bit confusing, given the automatic simplification uses the manual simplification.  Maybe, consider using "interactive simplification" as opposed to "manual simplification" and refer Algorithm 1 directly while discussing automatic simplification*.

We replaced the term manual simplification to interactive simplification as the new term is more descriptive and we also referred the Algorithm 1 directly from the description of automatic simplification. The changes have been made in Section Trajectory Simplification.

1. *The current formulation of trajectory complexity c(x) suggests that acute alpha's will lead to less complex trajectories, which isn't necessarily accurate. If c(x) is a curvature measure, then what you should care is the deviation of alpha from PI*.

We would like to thank you for this remark since the formulation and its description was misleading. We are interested in the value of alpha since an acute angle between directions of the two consecutive segments means that the direction of the trajectory is not changing. Therefore, the value of resulting complexity *c(x)* of segments with low alphas is low. On the other hand, an angle approaching PI means that the direction changed significantly and the new direction of the trajectory is close to opposite of its previous direction. We edited the description of complexity measure in Section Trajectory Simplification to better reflect this. We hope that the formulation is more clear now.

**Reviewer 4**

1. *While the proposed technique of simplification is interesting, it is not particularly novel and concerns with path simplification in such domains are not clearly addressed by the authors. The expert involvement is minimal and some more insights/comments would have been helpful.*

In this paper, our primary goal was to introduce a visual exploration approach rather than to propose a novel path simplification technique. Naturally every simplification technique introduces a risk of removing an important piece of data. This issue was discussed during the design stage and as such the proposed technique met the criteria specified by the biochemists. Additional explanation can be found in Comment 2.

1. *Some of the concerns that are not discussed in the paper are related to the simplification algorithm and the presentation of the simplified path. Based on our experience, domain experts are not particularly delighted about the use of simplification. The proposed manual simplification seems to simplify on top of the proposed automated simplification. There may be situations where an expert may want to see more fine grained details and constant simplification throughout may not be the answer.*

The reviewer is right that generally the domain experts are not fond of any simplification and prefer to explore the original data. However, in this case they admitted that some sort of simplification is necessary to get at least a rough idea about the ligand movement. The original data are too scattered to be able to understand the significant moments from it. Naturally, the user can anytime return to the original data so the simplification can be taken only as a guidance through the data. Moreover, by enabling the interactive simplification on selected parts of the trajectory, the user can adjust the trajectory according to his or her needs.

1. *With regards to path/trajectory/line simplification, there is a classic algorithm called the Douglas-Peucker algorithm. I would recommend that the authors take a look at the algorithm and the following improvements on the algorithm.*

*Visvalingam, Mahes, and J. Duncan Whyatt. "The Douglas‐Peucker Algorithm for Line Simplification: Re‐evaluation through Visualization." Computer Graphics Forum. Vol. 9. No. 3. Blackwell Publishing Ltd, 1990.*

*Hershberger, John Edward, and Jack Snoeyink. Speeding up the Douglas-Peucker line-simplification algorithm. University of British Columbia, Department of Computer Science, 1992.*

We employed the Douglas-Peucker algorithm within a non-smoothing simplification approach (see Comment 1). The non-smoothing simplification is described in the end of Section Automatic Simplification.

1. *The authors may find the following paper relevant to the visualization in Figure 7 interesting:*

*Rodgers, Peter, Gem Stapleton, and Peter Chapman. "Visualizing sets with linear diagrams." ACM Transactions on Computer-Human Interaction (TOCHI) 22.6 (2015): 27.*

We thank the reviewer for this valuable comment. The suggested paper postulates six questions regarding the linear diagrams visualization (and answers them via a user study). We went through the paper and updated our visualization accordingly. Most of the techniques mentioned in the paper were already presented in our visualization except two of them. We have added the guidelines to indicate the start and end of overlaps in our visualization since the paper proved that study participants performed significantly better when using these guidelines. On the other hand, we cannot lower the amount of segments since in our case the time dependency is more important. Hence we cannot change the segments placement in order to minimize their number.

1. *Overall the paper presents an interesting application of a particular path simplification algorithm to a problem. The domain expert interaction is very minimal. For a paper such as this which is so heavily focused on an application domain, it is critical to have a significant section of the paper discussing domain expert evaluation of the proposed technique. The paper currently has an "Analysis Procedure and Discussion" section, but it is not clear how much interaction the domain expert had and whether any deeper insights were found. Usually, it also helps to show that the technique was applicable in more than one specific case.*

To address this issue, we added the information about the cooperating group of domain experts and how they participated on the design and evaluation. Moreover, we added another scenario of possible usage of our tool – the second scenario describes more complex case when three ligands are entering a protein structure. The second scenario was included into Section Analysis Procedure and Discussion.